

Rapport - Prédiction de Structure Secondaire (Non achevé)

KEDDIS Adam

March 3, 2024

Contents

1	Introduction	3
1.1	Les protéines	3
1.2	Les AA	3
1.3	Structure des protéines	3
1.4	Hélice et Feuillet	3
2	Projet	3
2.1	Données	3
2.2	Les variables	3
2.3	L'analyse	3
3	Résultats	3
3.1	Les variables	4
3.2	Les corrélations	4
3.3	La normalité	4
3.4	L'ACP et les variables sélectionnées	4
3.5	Comparaison Hélice vs Feuillet	4
3.6	Clustering et K-means	4
3.7	Clustering et CAH	4
3.8	Classification et KNN, Homologie	4
3.9	Passage en qualitatif	4
3.10	Apprentissage supervisé et K-ID3	4
3.11	Mutation induite	4
3.12	Forêt	4
4	Conclusion et limite	5
4.1	Les paramètres	5
4.2	Les données	5
4.3	Conclusion	5

1 Introduction

Le calcul de la structure d'une protéine est un problème classique en bio-informatique. Aujourd'hui considéré comme un dogme: *La structure permet la fonction*. La compréhension du repliement, permettrait d'identifier la fonction d'une protéine à partir de sa séquence en acide aminé (AA à partir d'ici). Avec un regard posé sur l'avenir, même généré des séquences répondant aux structures désirées et donc à la fonction désirée. Cela permettrait par exemple la conception d'outil moléculaire très fins ou encore de médicament.

1.1 Les protéines

=> Transcription etc ...

1.2 Les AA

1.3 Structure des protéines

=> Les différents niveau ...

1.4 Hélice et Feuillet

=> AA etc

2 Projet

...

2.1 Données

...

2.2 Les variables

...

2.3 L'analyse

...

3 Résultats

...

3.1	Les variables
...	
3.2	Les corrélations
...	
3.3	La normalité
...	
3.4	L'ACP et les variables sélectionnées
...	
3.5	Comparaison Hélice vs Feuillet
...	
3.6	Clustering et K-means
...	
3.7	Clustering et CAH
...	
3.8	Classification et KNN, Homologie
...	
3.9	Passage en qualitatif
...	
3.10	Apprentissage supervisé et K-ID3
...	
3.11	Mutation induite
...	
3.12	Forêt
...	

4 Conclusion et limite

...

4.1 Les paramètres

=_i trop de possibilité

4.2 Les données

=_i Escherichia coli only sur le test

4.3 Conclusion

...